DOI:10.14075/j.jgg.2024.11.525

文章编号:1671-5942(2025)12-1294-10

一种基于 *K*-means 聚类的最小二乘配置法及在大尺度地壳速度场高精度拟合推估中的应用

王 程[1] 瞿 伟[1,2] 李久元[1] 唐兴友[1] 王光耀[1] 杨 凯[1]

1 长安大学地质工程与测绘学院,西安市雁塔路 126 号,710054

2 西部矿产资源与地质工程教育部重点实验室,西安市雁塔路 126 号,710054

摘 要:针对传统最小二乘配置(traditional least squares collocation,TLSC)算法在大尺度区域地壳运动速度场高精度拟合中,在监测站点稀疏区域与块体边缘处速度场拟合结果会出现异常与不平滑的问题,结合 *K*-means 聚类算法与 TLSC 算法发展了一种基于 *K*-means 聚类的最小二乘配置法(KLSC),并在青藏高原 GNSS 地壳运动实测速度场中验证该方法的有效性。结果表明:1)相较于 TLSC 算法,KLSC 算法利用 *K*-means 算法在无监督分类中的优势,基于 GNSS 速度场本身特征先将研究区域划分为多个速度相似的子区域,然后在每个子区域内分别利用 TLSC 进行速度场拟合,避免了局部复杂地质环境对区域速度场拟合精度的影响;2)KLSC 算法以各网格点到各聚类中心的距离最近为依据选取拟合参数,解决了数据稀疏区域速度场拟合结果较差的问题;3)KLSC算法利用次近距离拟合并结合卷积滤波,有效解决了块体边缘处速度场拟合结果不平滑的问题;4)KLSC算法拟合的速度场的 RMSE 精度和相关性均优于 TLSC 算法,东、北向拟合速度场 RMSE 精度分别提高 37%~48.2%和 52.1%~67.2%,相关性分别提高 24.1%~24.7%和 4.7%~5.2%。

关键词:GNSS 速度场;最小二乘配置算法;*K*-means 算法;青藏高原;地壳平滑速度场

中图分类号:P228;P313 **文献标识码**:A

地壳运动速度场的高精度拟合推估可以较好地表征出区域地壳运动形变特征,尤其是在大尺度复杂地质构造背景下,高精度、高覆盖的地壳运动速度场可以为揭示区域构造应力应变状态[1-2]、次级地块间相互作用[3]以及地震活动分析[4]等提供重要的基础依据。

在众多速度场拟合方法中,最小二乘配置(least squares collocation,LSC)算法由于其自身的优势被广泛应用[5-6]。但该方法在大尺度区域地壳运动速度场高精度拟合过程中还存在一些问题,如对统一运动趋势中的局部复杂变化区域表现较差,在监测站点稀疏区域速度场拟合结果会出现异常,以及在块体边缘处速度场拟合结果会出现不平滑的现象[7]。因此,本文利用 *K*-means 分类在无监督分类中的优势,结合 LSC 算法,发展了一种基于 *K*-means 的最小二乘配置算法(KLSC)。该算法首先通过 *K*-means 聚类,根据 GNSS 速度场的本身特征将研究区域划分为多个速度相似的子区域,然后在每个子区域内利用 TLSC 进行速度场拟合,从而有效提高局部复杂地质环境区域的拟合精度,解决传统方法在此类区域拟合精度较低的问题。在连续速度场插值过程中,进一步以各网格点到各聚类中心的距离最近为依据选取拟合参数,可解决监测站点稀疏区域速度场拟合结果较差的问题。最后利用次近距离拟合,即将超出各类插值速度中位数 3 倍标准差的网格点划归到次近距离类别中后重新插值,并结合卷积滤波解决块体边缘处速度场拟合结果不平滑的问题。

1 最小二乘配置改进模型

TLSC 算法是一种可以同时求定非随机性倾向参数和随机性信号的最优估计方法。将该方法用于建立 GNSS 观测地壳运动速度场模型[5-6,8]。

1.1 *K*-means 聚类算法及其最优解判定指标

K-means 聚类是一种非监督学习算法,主要

收稿日期:2024-11-18

项目来源:国家自然科学基金(42174006);陕西省杰出青年科学基金(2022JC-18);长安大学中央高校基本科研业务费专项(300102263201)。

第一作者简介:王程,硕士生,主要从事地壳形变大地测量反演研究,E-mail:2022126014@chd.edu.cn。

通讯作者:瞿伟,博士,教授,博士生导师,主要从事地壳形变大地测量反演、地质灾害高精度监测及成因机理研究,E-mail:maikerqq@163.com。

用于将数据集划分为 K 个不同的簇,其核心思想是通过迭代优化,使得每个簇中的数据点与簇中心之间的距离尽可能小,同时不同簇之间的中心距离尽可能大[9]。

由于 K-means 的聚类数目 K 可任意选取,初始聚类中心位置 C 也是随机确定的,会导致相同条件下每次聚类产生的效果不尽相同。因此,本文利用类内平方和(SSE)、轮廓系数、Calinski-Harabasz 指数[9]和 Gap 统计量[10]对 K-means 聚类的最优解进行评估确定。

肘部法则(elbow method)是一种常用的聚类算法评估方法,用于确定最佳聚类数 K。其核心思想是通过观察聚类数增加时 SSE 或其他性能指标的变化趋势,当 K 值较小时,增加 K 会显著降低误差,但超过某一临界点后,误差的下降速度趋于平缓,形成“肘部”形状,该临界点对应的 K 值被认为是较优的聚类数。肘部位置通过计算聚类数目与 SSE 曲线斜率变化最大的点来确定。

在利用肘部法则确定最佳 K 值后,可计算出当前 K 值时的轮廓系数、Calinski-Harabasz 指数、Gap 统计量值所占总值的比例 r:

$$r=\frac{K(m)-\min(m)}{\max(m)-\min(m)} \tag{1}$$

式中,$K(m)$为肘部法则确定最佳 K 值后需计算 r 的指标值,m 指代的指标可为轮廓系数、Calinski-Harabasz 指数、Gap 统计量值,$\max(m)$、$\min(m)$对应该指标的最大、最小值。

在实际应用中,通常绘制聚类数目与 SSE、平均轮廓系数、Calinski-Harabasz 指数和 Gap 统计量之间的曲线,利用肘部法则找到 SSE 的减少速度明显放缓,即聚类数目与 SSE 曲线斜率变化最大的点,同时后三者最大化时对应的聚类数目为最佳选择。

1.2　KLSC 算法

KLSC 算法的基本思路是,首先利用 K-means 聚类判定指标优选出最佳的聚类数 K,再利用 K-means 聚类算法将研究区域 GNSS 速度场按位置和东、北向分量进行聚类,每个类别再分别利用 TLSC 拟合推估出拟合速度场。进而在连续速度场插值过程中,以各网格点到各聚类中心的距离为依据,选取不同的拟合参数,利用次近距离拟合处理异常数据,并进一步运用卷积滤波解决块体边缘处速度场拟合结果不连续的问题。计算步骤概述如下:

1)利用上述准则综合判定 K-means 聚类多个准最佳聚类数目 K,并判定当前 K 值下聚类中心的稳定性。

2)将观测速度场中的拟合点按指定 K 值进行分类,利用 TLSC 计算每个类别的拟合参数并拟合观测速度场。

3)根据 KLSC 算法多次拟合结果的 RMSE、相关系数等指标判定拟合点与检核点拟合结果的稳定性,并与 TLSC 算法结果进行对比分析,最后综合选取精度更高、相关性更好、结果更稳定的 K 值作为最佳聚类数目,进行后续实验。

4)将目标分辨率的网格点按照到各聚类中心的最近距离划分归属类别,并利用相关拟合参数进行插值,获取连续速度场。

5)获取连续速度场后,若出现异常值,则采用次近距离法对其处理,计算拟合东、北向速度分量中位数 M 和标准差 σ。以$[M-3\sigma,\ M+3\sigma]$为界,超出此边界范围的值视作异常值,异常值格网点使用与其距离次近的类别的拟合参数进行拟合,以此类推,直至所有类别均满足条件后输出结果。

6)在各类数据均无异常后,若各类别边缘速度有明显的不连续现象,再采用卷积滤波插值东、北向速度分量。针对边缘区域采用加权滤波核,例如对于 3×3 的均值滤波卷积核为[1 1 1; 1 1 1; 1 1 1]/9,分别将卷积核对应的边缘元素设为 0,如上侧边缘的卷积核为[0 0 0; 1 1 1; 1 1 1]/6,以此类推,最终获取研究区域平滑的连续速度场。

KLSC 算法技术路线如图 1 所示。

2　实例测试

本文采用 Wang 等[11]公布的青藏高原 75°～105°E、25°～40°N 区域的 557 个 GNSS 速度场数据,并从中随机选取超总数 10%且均匀分布的 60 个点作为检核点,其余点则用于构建速度场拟合推估模型(图 2)。

2.1　KLSC 聚类数目确定与稳定性分析

K-means 聚类较为依赖聚类数目的选择和聚类中心的确定,不同的聚类数目和聚类中心往往会对聚类结果产生较显著的影响[12]。但由于 K-means 方法中聚类数目的确定较为困难,通常需要借助肘部法则或轮廓系数法等方法辅助判断,同时初始聚类中心是随机选取的,不同的初始中心可能导致算法结果存在较大差异,多次实验可能得出不同的聚类结果。因此,本节着重分析聚类数目 K 值的选择和聚类中心的稳定性。

2.1.1　聚类数目 K 的确定

本文取初始簇质心位置重复聚类 100 次时 SSE最小值对应的聚类结果作为最终结果(该方

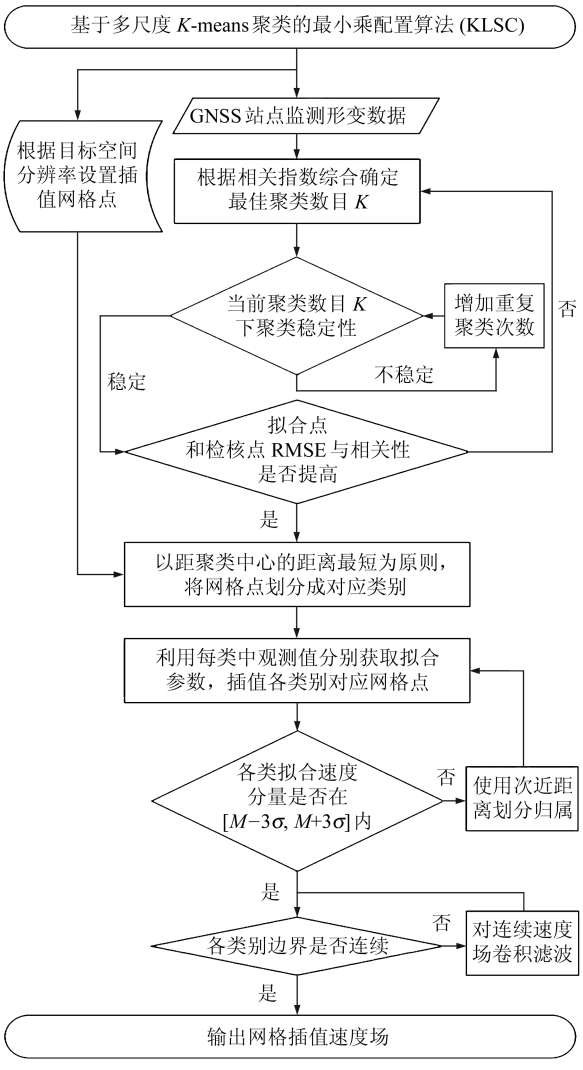


图 1 KLSC 算法流程

Fig. 1 Flow chart of KLSC algorithm

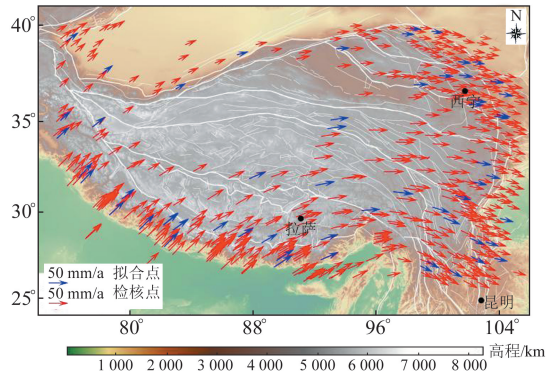


图 2 青藏高原区域 GNSS 地壳运动观测速度场

Fig. 2 GNSS crustal movement observation velocity field in the Qinghai-Xizang plateau region

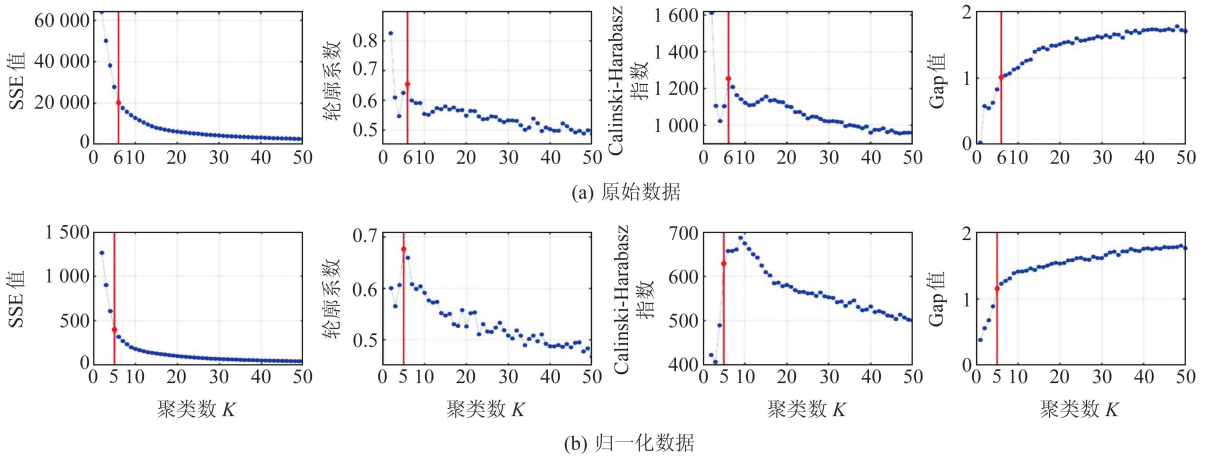
法简称为 100 次聚类方法)，设置最大 K 值为 50，用以判断最佳聚类中心数 K。

K-means 聚类可以考虑将数据进行归一化，以消除不同量纲和数量级对 SSE 等指标的影响。K-means 聚类时，以观测速度场的经度、纬度、速度东向分量和速度北向分量 4 个特征为依据，利用极差和标准差公式获取的观测数据特征范围和特征标准差如表 1 所示。从表中可以看出，所有数据特征的范围和标准差均在同一数量级，但速度北向分量极差较大，因此分别利用归一化前与归一化后数据进行聚类，判断最佳 K 值(图 3)。

表 1 原始数据特征

Tab. 1 Characteristics of the original data

	经度 /(°)	纬度 /(°)	速度东向分量 /mm·a^{-1}	速度北向分量 /mm·a^{-1}
特征范围	31.905	13.953	33.160	60.820
特征标准差	10.290	4.274	6.157	17.644



(a) 原始数据

(b) 归一化数据

图 3 原始数据归一化前后聚类数目与各聚类指标的关系

Fig. 3 Relationship between the number of clusters and each clustering index before and after normalization of original data

从图 3(a)可以看出，对于原始数据，利用肘部法则确定出的最佳聚类数目均为 6，且此时轮廓系数(占比 0.79)、Calinski-Harabasz 指数(占比 0.78)和 Gap 统计量(占比 0.56)的值也较大。由图 3(b)可以看出，数据归一化后，利用肘部法则确定出的最佳聚类数目为 5，且此时轮廓系数(占比 1.00)、Calinski-Harabasz 指数(占比 0.92)和 Gap 统计量(占比 0.54)的值也较大。此外，当

K=9 时,4 个指数也基本处于合理值范围内。

可以看出,归一化对最佳聚类数目几乎无影响。为了充分考虑 GNSS 真实数据的特征,后续实验不对数据进行归一化。同时,为了获取最佳聚类结果,分别考虑聚类数目 K=5、6、9 进行实验,并结合插值效果进行判断。

2.1.2　聚类中心分布稳定性测试

确定聚类数目 K 值后,需要判定当前 K 值下的聚类稳定性。利用观测速度场中的拟合点,取 100 次聚类方法结果作为最终结果。分别得出 K=5、6、9 时的速度场分类情况,如图 4 所示。

图 4 中黑点为聚类中心所在位置,聚类数目 K=5、6、9 时,最佳 SSE 值分别为 400.340、319.462、201.659;平均 SSE 值分别为 400.358、319.465、201.683;平均轮廓系数分别为 0.677、0.660、0.604;平均 Calinski-Harabasz 指数分别为 628.593、656.973、686.865。按 100 次聚类方法实验 100 次,获取 SSE 最小值(即最佳 SSE)、SSE 均值、轮廓系数均值和 Calinski-Harabasz 指数均值作为参考,验证 100 次聚类方法的稳定性(图 5)。

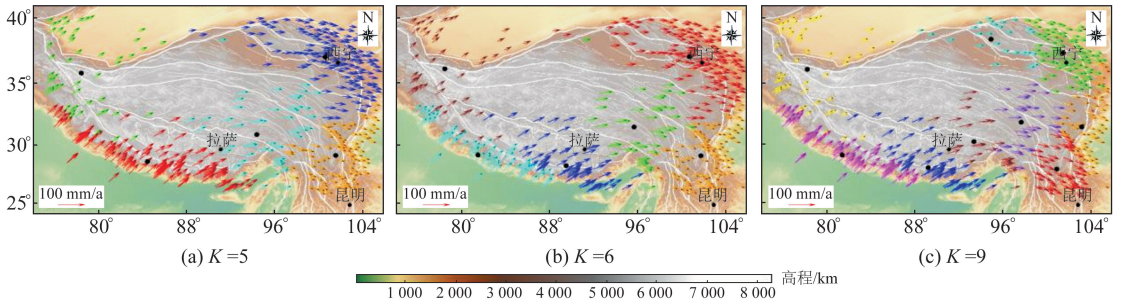


图 4　基于 KLSC 算法的速度场聚类结果

Fig. 4　Velocity field clustering results based on KLSC algorithm

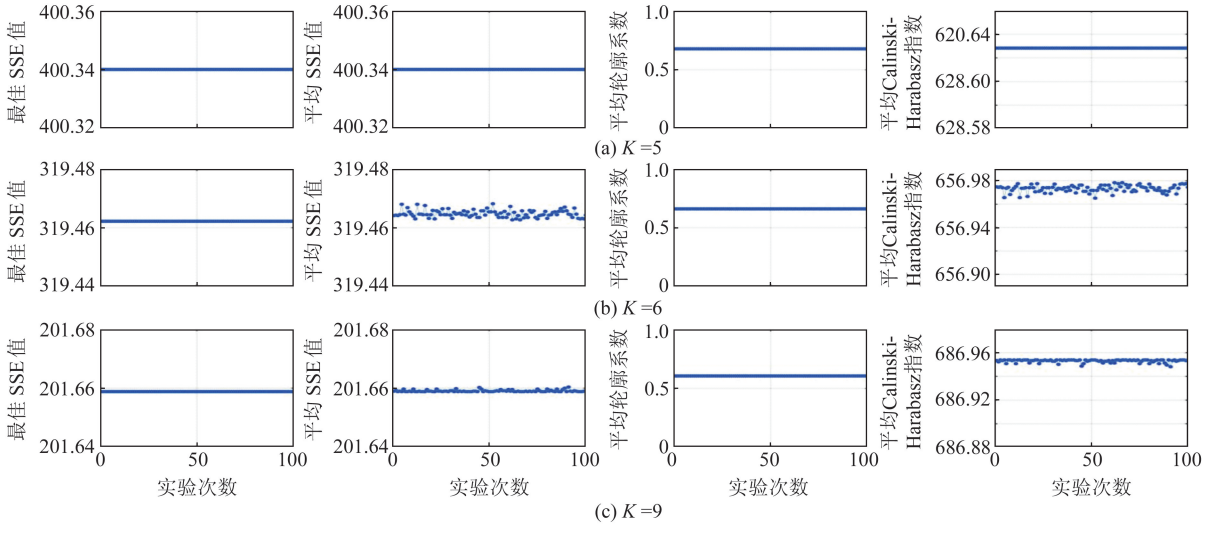


图 5　100 次聚类方法指标分布

Fig. 5　Indexes distribution of 100 times clustering method

图 5 显示,100 次聚类最佳 SSE 值不会发生改变,平均 SSE 值、平均轮廓系数、平均 Calinski-Harabasz 指数也基本不变,具有很强的平稳性,且均可以稳定找到最优聚类中心。

2.2　区域地壳运动速度场拟合推估结果与对比

在聚类数目 K=5、6、9 时,分别利用 KLSC 与 TLSC 算法获得青藏高原及周边区域地壳运动模拟与残差速度场(图 6)。

由图 6(a)看出,基于 TLSC 算法获取的青藏高原整体运动呈现出 NE-SEE 向顺时针旋转运动趋势。从图 6(e)看出,TLSC 算法在喜马拉雅东构造结区域模拟效果较差,在其余区域模拟效果较好。相较于图 6(a),图 6(c)、6(e)、6(g)基于 KLSC 算法获取的拟合速度场在青藏高原东南缘出现较明显的变化。相较于图 6(b),从图 6(d)、6(f)和 6(h)发现,基于 KLSC 算法获取的残差速度场量级显著减小,特别是在喜马拉雅东构造结区域模拟速度场与实测速度场符合度更高。

3　精度对比分析

为进一步分析 TLSC 与 KLSC 算法的拟合精度,计算 2 种算法的拟合点和检核点的 RMSE 和相关系数(表 2、图 7),同时绘制各点残差的分布情况(图 8)。

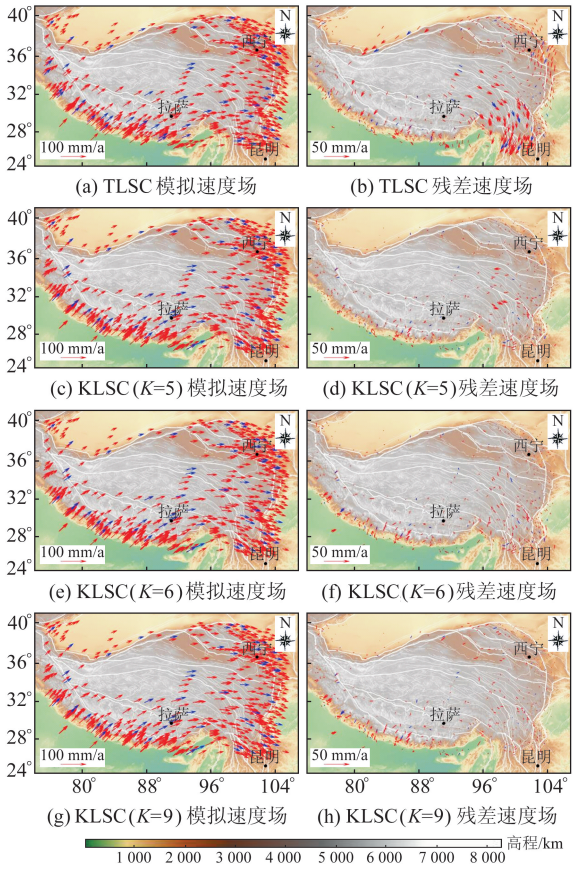


图 6 基于 TLSC 与 KLSC 算法的研究区域观测、模拟与残差速度场

Fig. 6 Observed, simulated and residual velocity fields based on TLSC and KLSC algorithms

表 2 基于 TLSC 和 KLSC 算法的拟合点和检核点的 RMSE 和相关系数

Tab. 2 The RMSE and correlation coefficient of fitted points and check points based on TLSC and KLSC algorithms

算法	拟合点 RMSE /mm·a^{-1}		检核点 RMSE /mm·a^{-1}		相关系数 R	
	东向	北向	东向	北向	东向	北向
TLSC	3.75	5.86	4.07	6.20	0.730	0.945
KLSC(K=5)	2.36	1.91	2.11	2.11	0.906	0.994
KLSC(K=6)	2.29	2.66	2.17	3.01	0.909	0.989
KLSC(K=9)	2.21	2.09	2.14	2.48	0.918	0.993

3.1 拟合结果精度分析

图 7 中红色箭头表示 KLSC 算法相对于 TLSC 算法拟合与推估的精度提升率，箭头下方数字表示相应的提升百分比。由图 7 可以看出，KLSC 算法相较于 TLSC 算法的 RMSE 精度均有大幅提升，东向与北向 RMSE 精度分别提升约 37%～48.2%、52.1%～67.2%。从表 2 中也可以看出，KLSC 算法拟合点与核验点的东、北向分量相关系数显著高于 TLSC 算法，平均分别提高约 24%、5%(表 2 与图 7)。

K=6 时的精度和相关性整体略低于 K=5、9 时，这是因为 K=6 作为聚类数目划分情况与

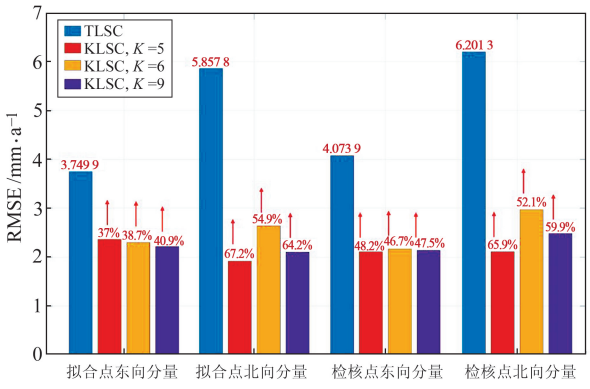


图 7 TLSC 与 KLSC 算法拟合点和检核点 RMSE 对比

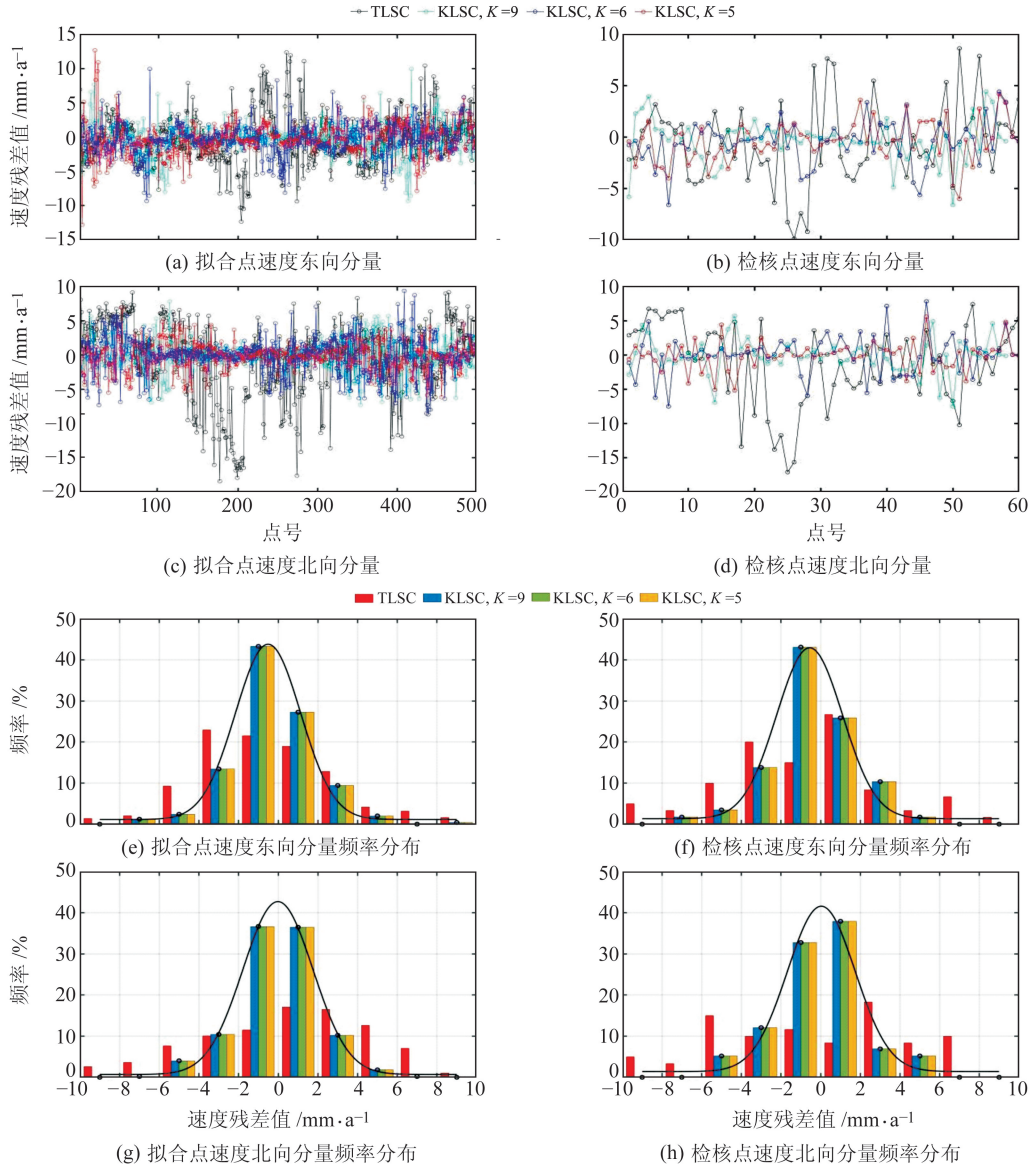
Fig. 7 Comparison of RMSE of fitted points and check points of TLSC and KLSC algorithms

青藏高原速度场实际情况不符。总体而言，K=5 时计算效果最佳。

从图 8 可以看出，KLSC 算法相较于 TLSC 算法，残差分布更集中在 0 附近，残差较小值出现频率大幅提高，更加符合正态分布。K=5、6、9 的残差出现频率没有太大变化。

如表 3 所示，在显著性水平 α=0.01 下，检核点的残差数据通过了 t 检验(P>0.01)，表明其符合正态分布假设，模型拟合效果良好，误差分布合理。然而，在拟合点残差方面，仅 KLSC 算法的北向残差通过了 t 检验(P>0.01)，而东向残差和 TLSC 算法的北向残差未能通过 t 检验(P<0.01)。分析原因发现，拟合过程中个别点出现较大残差(图 8(a)、8(c))，显著影响了整体残差的均值和方差，导致其偏离正态分布。去除残差中的 1 组极值后重新进行 t 检验，可以发现，除 TLSC 算法因未能更好地捕捉局部复杂变化，以及 K=9 时块体划分不合理分别导致北向和东向残差未能通过 t 检验外，其余均通过了 t 检验(P>0.01)。这表明模型在大多数情况下满足正态分布假设，拟合效果较好，误差分布合理。

将 KLSC 算法与以下几种方法进行对比：多尺度球面小波[13](中误差为 3.2 mm/a，残差值主要集中在－5～＋5 mm/a 之内)、球面最小二乘配置方法[8](其东、北向 RMSE 分别为 4.61 mm/a、6.85 mm/a)、总体最小二乘配置算法[6](检核点 RMSE 约为 2.4 mm/a)、顾及距离尺度因子的自适应最小二乘配置融合新算法(ADLSC)[14](检核点 RMSE 约为 2.5 mm/a)、分块体欧拉矢量模型[15](西藏块体的东、北向 RMSE 分别为 6.13 mm/a、7.07 mm/a)。虽然上述对比的研究为青藏高原内的部分区域，其研究区域的地壳运动复杂程度相较于整个青藏高原而



(a) 拟合点速度东向分量　(b) 检核点速度东向分量

(c) 拟合点速度北向分量　(d) 检核点速度北向分量

(e) 拟合点速度东向分量频率分布　(f) 检核点速度东向分量频率分布

(g) 拟合点速度北向分量频率分布　(h) 检核点速度北向分量频率分布

图 8　TLSC 与 KLSC 算法残差分布与频率分布直方图

Fig. 8　Residuals distribution and frequency distribution histograms of TLSC and KLSC algorithms

表 3　基于 TLSC 与 KLSC 算法的拟合点和检核点残差 t 检验结果

Tab. 3　The t-test results of the residuals of fitting points and check points based on TLSC and KLSC algorithms

数据	算法	拟合点残差 t 值		拟合点残差 P 值		检核点残差 t 值		检核点残差 P 值	
		东向	北向	东向	北向	东向	北向	东向	北向
原始残差	TLSC	−2.600	−2.815	0.010	0.005	−0.977	−1.835	0.333	0.072
	KLSC(K=5)	−2.752	−1.608	0.006	0.109	−1.412	−0.423	0.163	0.674
	KLSC(K=6)	−2.652	−0.711	0.008	0.478	−1.566	−0.008	0.123	0.994
	KLSC(K=9)	−3.062	−0.590	0.002	0.556	−1.637	−0.357	0.107	0.722
去除极值后数据	TLSC	−2.477	−2.698	0.014	0.007	−0.696	−1.564	0.489	0.123
	KLSC(K=5)	−2.583	−1.479	0.010	0.140	−1.109	−0.110	0.272	0.913
	KLSC(K=6)	−2.509	−0.572	0.012	0.568	−1.258	0.341	0.214	0.735
	KLSC(K=9)	−2.922	−2.922	0.004	0.661	−1.336	0.035	0.187	0.973

言也相对较弱,但从本文方法检核点 RMSE 与这些方法的对比来看,也进一步验证了本文提出的 KLSC 方法在速度场拟合中的优越性与适用性。

3.2　多次实验判定结果稳定性

进行单次实验不能完全证明 KLSC 算法优于 TLSC 算法,故进一步开展以下多组实验。考虑同一组检核点实验 50 次精度分布问题,并绘制每次实验的 RMSE 分布图(图 9)。

从图 9 看出,当 K=5、6 时,重复实验 50 次,利用 100 次聚类方法基于 KSLC 算法获取的拟合点与检核点的 RMSE 不会出现波动,并不会对结

果产生影响。但当 $K=9$ 时，利用 100 次聚类方法 KLSC 算法获取的拟合点与检核点的个别 RMSE 会出现明显波动，且东向速度误差波动较大，极差超过 2 mm/a，北向误差极差在 1 mm/a 左右。$K=9$ 时稳定性较差，分析其原因是块体划分较多，K-means 聚类中心的随机性导致边界点归属容易出现不稳定。

经过在精度、相关性、残差分布和稳定性等方面的综合比较与分析，最终选择 $K=5$ 作为后续实验的聚类数。

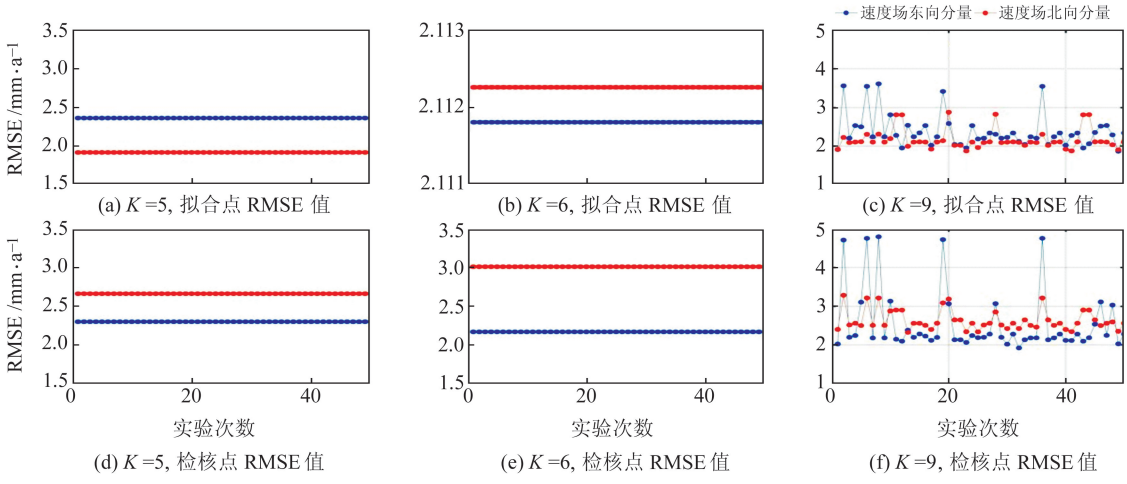


图 9 重复实验 50 次 100 次聚类方法拟合点与检核点总残差变化

Fig. 9 Variations of total residuals of fitting points and check points of 100 times clustering method after repeating the experiment 50 times

4 连续平滑地壳运动速度场获取

在确保 KLSC 算法的稳定性后，为获取区域连续的平滑速度场，更好地分析区域整体地壳形变特征，需要对区域地壳运动速度场进行网格速度场插值。图 10 和图 11 分别为利用 TLSC 算法和 KLSC 算法获取的青藏高原地区 GNSS 连续速度场。

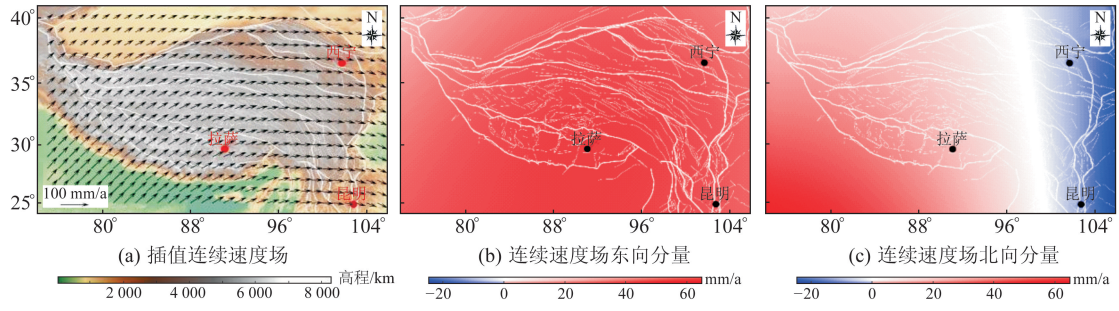


图 10 基于 TLSC 算法插值的间隔 1°的地壳连续速度场及其分量

Fig. 10 The crustal continuous velocity field interpolated at intervals of 1° based on the TLSC algorithm

为解决监测数据稀疏区域地壳运动速度场拟合结果较差的问题，在连续速度场插值过程中以各网格点到各聚类中心的距离为依据选取拟合参数；然后利用次近距离拟合处理异常数据；最后利用卷积滤波解决边界不连续的问题，确保速度场在块体边缘的平滑性和连续性。为评价滤波效果，计算各滤波前后均方误差(MSE)和滤波后速度场分量的梯度平方和(表 4)，并绘制基于 KLSC 算法插值连续速度场的速度东、北向分量图(图 12)。

对比图 10(a)与图 11 滤波后速度场可以看出，相较于 TLSC 算法，基于 KLSC 算法获取的青藏高原网格连续速度场整体更符合青藏高原实际地壳运动趋势，特别是在喜马拉雅东构造结区域。如图 11(a)所示，基于 KLSC 算法的连续速度场拟合推估结果中会出现部分网格点位置处的插值速度异常。这可能是采用最近距离法划归类别导致的，采用次近距离法划分可避免因类别归属错误而出现异常值的情况(图 11(b))。

表 4 各滤波结果指标

Tab. 4 Indices of each filtering results

滤波方法	滤波前后 MSE /mm·a^{-1}		滤波后速度场(梯度平方和)/a^{-2}	
	东向	北向	东向	北向
均值滤波	4.253	7.079	5.329	2.947
高斯滤波	5.945	9.004	4.618	2.210
中值滤波	1.859	6.312	7.614	4.180
双边滤波	0.011	0.015	11.072	12.382
核密度估计法滤波	0.037	0.083	11.170	12.309

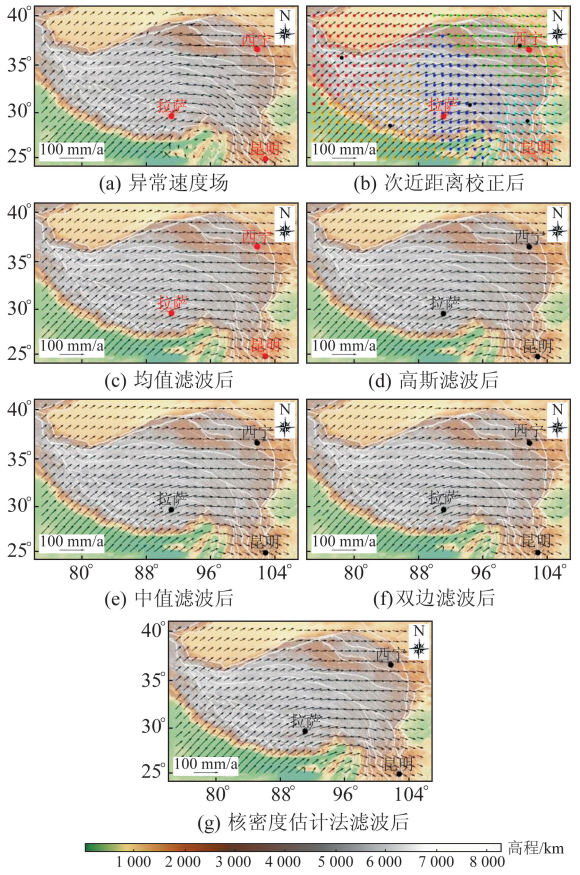


图 11　基于 KLSC 算法插值的间隔 1°的地壳连续速度场

Fig. 11　The crustal continuous velocity field interpolated at intervals of 1° based on the KLSC algorithm

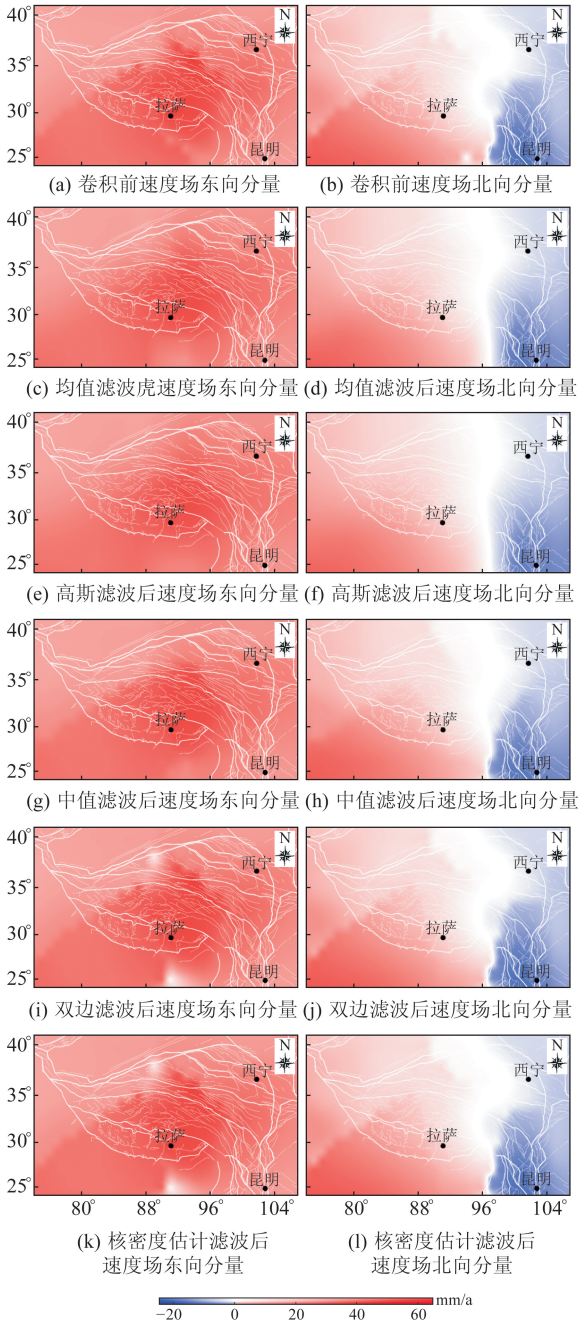


图 12　基于 KLSC 算法插值的间隔 1°的地壳连续速度场分量

Fig. 12　The crustal continuous velocity field components interpolated at intervals of 1° based on KLSC algorithm

从图 11(b)也可以发现,块体边缘处速度呈现出明显的不连续现象。经过滤波后,卷积前后速度场会在块体边缘处发生较明显的突变。由图 11(c)～11(g)的 5 种滤波后速度场与图 12 的滤波后速度场东、北向分量可见,高斯滤波与均值滤波差别不大,效果均较好,其余 3 种滤波边缘去除效果较差。通过表 4 可以看出,高斯滤波前后 MSE 最大,梯度平方和最小,说明高斯滤波使滤波前后的边缘消除效果最好,在保持原有运动趋势的同时,拟合速度场在块体边缘处呈现出较好的平滑性和连续性,且更符合青藏高原地区实际地壳运动变形特征。

5　结　语

1)与 TLSC 算法不同,KLSC 算法并非直接对整个大尺度区域进行速度场拟合,而是先通过 K-means 聚类将区域划分为若干子区域,并在每个子区域内分别应用 TLSC 进行速度场拟合,从而有效避免局部复杂地质环境对拟合精度的影响。

2)在连续速度场插值过程中,KLSC 算法依据各网格点到聚类中心的最小距离选取拟合参数,显著改善了数据稀疏区域速度场拟合结果较差的问题。此外,通过引入次近距离拟合并结合卷积滤波,成功解决了块体边缘处速度场拟合结果不平滑的难题。

未来的研究工作将重点放在进一步优化 KLSC 方法及其在更大尺度和更复杂地质区域的推广应用上。针对 K-means 聚类的敏感性和卷积核设置的影响,进一步探索鲁棒性更强的聚类算法及自适应卷积核优化策略,深入拓展 KLSC 方法的泛化能力。

参考文献

[1] Li Y H, Liu M, Wang Q L, et al. Present-Day Crustal Deformation and Strain Transfer in Northeastern Tibetan Plateau[J]. Earth and Planetary Science Letters, 2018, 487: 179-189

[2] 瞿伟，高源，陈海禄，等. 利用 GPS 高精度监测数据开展青藏高原现今地壳运动与形变特征研究进展[J]. 地球科学与环境学报，2021，43(1)：182-204(Qu Wei, Gao Yuan, Chen Hailu, et al. Review on Characteristics of Present Crustal Tectonic Movement and Deformation in Qinghai-Xizang Plateau, China Using GPS High Precision Monitoring Data[J]. Journal of Earth Sciences and Environment, 2021, 43(1): 182-204)

[3] Hao M, Li Y H, Zhuang W Q. Crustal Movement and Strain Distribution in East Asia Revealed by GPS Observations[J]. Scientific Reports, 2019, 9

[4] 苏小宁，熊仁伟，孟国杰，等. 2021 年青海玛多 M_W7.3 地震地壳变形及区域孕震环境[J]. 地球物理学报，2022，65(3)：1032-1043(Su Xiaoning, Xiong Renwei, Meng Guojie, et al. The Characteristics of Crustal Deformation and Regional Seismogenic Environment Associated with the 2021 Madoi, Qinghai M_W7.3 Earthquake[J]. Chinese Journal of Geophysics, 2022, 65(3): 1032-1043)

[5] 王乐洋，陈汉清. 多波束测深数据处理的抗差最小二乘配置迭代解法[J]. 测绘学报，2017，46(5)：658-665(Wang Leyang, Chen Hanqing. Multi-Beam Bathymetry Data Processing Using Iterative Algorithm of Robust Least Squares Collocation[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(5): 658-665)

[6] 王乐洋，陈汉清，温扬茂. 地壳形变分析的总体最小二乘配置方法[J]. 大地测量与地球动力学，2017，37(2)：163-168(Wang Leyang, Chen Hanqing, Wen Yangmao. Analysis of Crust Deformation Based on Total Least Squares Collocation[J]. Journal of Geodesy and Geodynamics, 2017, 37(2): 163-168)

[7] 金阳，金红林，高原，等. 基于 GNSS 数据的三种应变计算方法在青藏高原东南缘地区的应用[J]. 地震，2021，41(4)：57-67(Jin Yang, Jin Honglin, Gao Yuan, et al. Application of Three Strain Calculation Methods Based on GNSS Data in the Southeastern Margin of Tibet Plateau[J]. Earthquake, 2021, 41(4): 57-67)

[8] 王栋，赵旭坤，刘望明. 基于最小二乘配置的板块边界地壳形变应变特征分析[J]. 华南地震，2021，41(3)：122-131(Wang Dong, Zhao Xukun, Liu Wangming. Characteristic Analysis of Crustal Deformation and Strain at Plate Boundary Based on Least Squares Configuration[J]. South China Journal of Seismology, 2021, 41(3): 122-131)

[9] 朱彬，杨诚，刘岩. GNSS 多径信号 3 种非监督学习法分析与比较[J]. 测绘学报，2021，50(12)：1762-1771(Zhu Bin, Yang Cheng, Liu Yan. Analysis and Comparison of Three Unsupervised Learning Clustering Methods for GNSS Multipath Signals[J]. Acta Geodaetica et Cartographica Sinica, 2021, 50(12): 1762-1771)

[10] 唐丹，张正军，王俐莉. 基于改进的近邻传播聚类算法的 Gap 统计研究[J]. 计算机技术与发展，2017，27(1)：182-185(Tang Dan, Zhang Zhengjun, Wang Lili. Study on Gap Statistic Based on Modified Affinity Propagation Clustering[J]. Computer Technology and Development, 2017, 27(1): 182-185)

[11] Wang M, Shen Z K. Present-Day Crustal Deformation of Continental China Derived from GPS and Its Tectonic Implications[J]. Journal of Geophysical Research: Solid Earth, 2020, 125(2)

[12] 任营营，王虎，王解先，等. 基于 K-Means++的省内子块体划分及中国大陆水平相对运动速度场模型的建立与分析[J]. 地球物理学报，2020，63(7)：2516-2533(Ren Yingying, Wang Hu, Wang Jiexian, et al. The Sub-Block Demarcation with K-Means++ in Each Province's Interior and Establishment Analysis of the Relative Horizontal Velocity Field Model in China's Mainland[J]. Chinese Journal of Geophysics, 2020, 63(7): 2516-2533)

[13] 胡良晨. 利用克里金插值研究中国大陆地壳变形特征[J]. 科学技术创新，2018(13)：10-11(Hu Liangchen. Study on Crustal Deformation Characteristics of Chinese Mainland by Kriging Interpolation[J]. Scientific and Technological Innovation, 2018(13): 10-11)

[14] Qu W, Chen H L, Liang S C, et al. Adaptive Least-Squares Collocation Algorithm Considering Distance Scale Factor for GPS Crustal Velocity Field Fitting and Estimation[J]. Remote Sensing, 2019, 11(22)

[15] 任雅奇. 基于 GPS 数据的中国地壳运动速度场模型的建立[D]. 郑州：信息工程大学，2012(Ren Yaqi. Establishment of Chinese Crustal Movement Velocity Model Based on the GPS Data[D]. Zhengzhou: Information Engineering University, 2012)

A *K*-Means-Based Least Squares Collocation Method and Its Application in High-Precision Fitting and Estimation of Large-Scale Crustal Velocity Fields

WANG Cheng[1] *QU Wei*[1,2] *LI Jiuyuan*[1] *TANG Xingyou*[1] *WANG Guangyao*[1] *YANG Kai*[1]

1 School of Geological Engineering and Geomatics, Chang'an University, 126 Yanta Road, Xi'an 710054, China

2 Key Laboratory of Western China's Mineral Resources and Geological Engineering, Ministry of Education, 126 Yanta Road, Xi'an 710054, China

Abstract: To address the issues of abnormal and non-smooth fitting results in sparse monitoring station

areas and block margins when using the traditional least squares collocation (TLSC) algorithm for high-precision fitting of crustal motion velocity fields in large-scale regions, a K-means clustering-based least squares collocation method (KLSC) is developed by integrating the K-means clustering algorithm with the TLSC algorithm. The effectiveness of this method is validated using the GNSS-measured crustal motion velocity field on the Qinghai-Xizang plateau. The results show that: 1) Compared to the TLSC algorithm, the KLSC algorithm leverages the advantages of the K-means algorithm in unsupervised classification to first divide the study area into multiple sub-regions with similar velocities based on the inherent characteristics of the GNSS velocity field, and then applies TLSC for velocity field fitting within each sub-region, thereby avoiding the impact of local complex geological environments on the accuracy of regional velocity field fitting; 2) The KLSC algorithm selects fitting parameters based on the proximity of each grid point to the cluster centers, resolving the issue of poor fitting results in data-sparse areas; 3) By utilizing secondary nearest-neighbor fitting combined with convolution filtering, the KLSC algorithm effectively addresses the non-smooth fitting results at block margins; 4) The RMSE accuracy and correlation of the velocity fields fitted by the KLSC algorithm are superior to those of the TLSC algorithm, with improvements in RMSE accuracy of 37% to 48.2% and 52.1% to 67.2%, and increases in correlation of 24.1% to 24.7% and 4.7% to 5.2% for the eastward and northward fitted velocity fields, respectively.

Key words: GNSS velocity field; least squares collocation method; K-means algorithm; Qinghai-Xizang plateau; smoothed crustal velocity field

Foundation support: National Natural Science Foundation of China, No. 42174006; Science Fund for Distinguished Young Scholars of Shaanxi Province, No. 2022JC-18; Fundamental Research Funds for the Central Universities, CHD, No. 300102263201.

About the first author: WANG Cheng, postgraduate, majors in geodetic inversion of crustal deformation, E-mail: 2022126014@chd.edu.cn.

Corresponding author: QU Wei, PhD, professor, PhD supervisor, majors in geodetic inversion of crustal deformation, geological disaster high-precision geodetic monitoring and disaster mechanism, E-mail: maikerqq@163.com.

(上接第 1271 页)

(LOS) coseismic deformation field of the M_S6.8 earthquake in Dingri, Xizang, on January 7, 2025. The geometric parameters of the seismogenic fault were derived through a Bayesian approach. Fault slip distribution was inverted using the SDM, and coseismic Coulomb stress changes were calculated. The inversion results indicate that the seismogenic fault has a strike of 187.34°, a dip angle of 56.40°, with an epicenter located at 87.50°E and 28.64°N, at a depth of 6.81 km. The maximum slip in the coseismic slip distribution is 4.24 m, with an average rake angle of −58.82°, suggesting that the earthquake was predominantly a normal faulting event with a minor component of left-lateral strike-slip. The inverted moment magnitude is M_W7.0, and the seismogenic fault is preliminarily identified as the Dengmo Co fault. Additionally, regions along the Zanda-Lhazê-Qiongduojiang fault, Darjeeling-Ngamring-Rinbung fault, and parts of the south Xizang detachment system fault, as well as the southern sections of the Xainza-Dingjie fault and Dengmo Co fault, exhibit Coulomb stress increases exceeding 0.1 bar, indicating that these areas warrant greater attention for future seismic hazards.

Key words: Dingri earthquake; coseismic deformation field; fault slip distribution; D-InSAR; Coulomb stress changes

Foundation support: Scientific Research Fund of Institute of Seismology, CEA and National Institute of Natural Hazards, MEM, No. IS202436359; Natural Science Foundation of Hubei Province, No. 2025AFB703.

About the first author: LI Chengtao, assistant researcher, majors in InSAR and GPS data processing and crustal deformation, E-mail: lict@cgps.ac.cn.

Corresponding author: LI Xue, associate researcher, majors in geodesy and geodynamics, E-mail: leexue1211@126.com.